Research

Random Forest Algorithms for County-level Supplemental Nutrition Assistance Program (SNAP) Classifications

Jaehyun Ahn
Texas A&M University
Agriculture and Life Sciences Building
College Station, TX 77843
(979) 985-1658
jaehyun.ahn@ag.tamu.edu

Matt Baker
Texas A&M University
Agriculture and Life Sciences Building
College Station, TX 77843
(979) 845-6944
mathew.baker@ag.tamu.edu

Jean Parrella
Texas A&M University
Agriculture and Life Sciences Building
College Station, TX 77843
(979) 845-2954
jean.parrella@ag.tamu.edu

Dawn Burton
Prairie View A&M University
761 Carden-Waller Cooperative Extension Building
Prairie View, TX 77446
(713) 797-7032
deburton@pvamu.edu

Rafael Quijada Landaverde
Texas A&M University
Agriculture and Life Sciences Building
College Station, TX 77843
(979) 458-7993
rafael.q@ag.tamu.edu

Random Forest Algorithms for County-level Supplemental Nutrition Assistance Program (SNAP) Classifications

Introduction/Need for Research

Since 1933, the Food Stamp Program has steadfastly supported American farmers/food suppliers and families/food consumers. The Program started amid The Great Depression for legislators to support federal government-level purchases of farm commodities and distribute those commodities to communities in need. In 2008, the name of the Program changed to the Supplemental Nutrition Assistance Program (SNAP), and its focus shifted to helping qualified consumers purchase fresh, nutritional foods (Snap to Health, n.d.). As of 2022, over 47 million U.S. consumers participated in SNAP. However, SNAP research in the discipline is limited, although the American Association of Agricultural Education Research Agenda underscores this (Food & Nutrition Service, 2023; Roberts et al., 2016). Considering agricultural educators and Extension professionals' essential roles in supporting an affordable, nutritious food supply and demand, research on SNAP is paramount. Understanding the needs of participants is essential for Program performance and evaluation (Pinard et al., 2017). We sought to classify counties receiving SNAP benefits based on variables that best-predicted participation.

Conceptual Framework

Identifying SNAP benefit inclusion and exclusion criteria is important as more households receive SNAP benefits. Annual data from the U.S. Department of Agriculture’s (USDA’s) Food Nutrition Service provides the most detailed demographic and geographic features regarding county-level SNAP participation (Newman & Scherpf, 2013). The purpose of the USDA collecting this data was to

1. Understand local needs affecting food insecurity to inform tailored SNAP implementation;
2. Help Extension professionals, policymakers, and administrators allocate resources effectively; and
3. Evaluate/classify counties with SNAP benefits if the Program meets its goals.

Methodology

The Data Planet (2023) is SAGE Publication’s web-based data repository containing over 65 billion data points, 550 databases, 90 data providers, and 16 subjects (e.g., Agriculture and Food). We accessed the repository and collected 86 SNAP-related, county-level quantitative variables predicting SNAP participation. The variables represented data from 3,143 U.S. counties between 2015 and 2017. For instance, we gathered information on the Food Environment (e.g., cluster of access to stores, food assistance, health activity, food price), SNAP Benefits Recipients (in percent), and socioeconomic characteristics originating from the Economic Research Service (2021), the Food and Nutrition Service (2023), and the United States Census Bureau (2022). We analyzed the data using TIBCO Statistica® (v. 13.0.5) for random forest classification.

Random forest is a decision tree-based machine learning (ML) method for classification, and its use enabled us to determine the correct classification of counties and predictor importance. Today, agriculture is one of the areas where ML is often applied (Pugliese et al., 2021). The random forest algorithm takes original data and creates many decision trees randomly in training. The test/trained data (i.e., data from the dataset that helps ML confirm the learned patterns) show the results from the combined decision trees that are less correlated. It is called an ensemble learning method in boosting-bagging-stacking that reduces bias and variance. All the predictors have ranked concerning the target variable (Ahn et al., 2022).

Findings

We bifurcated the target variable (i.e., 3,143 counties) according to the median of 2017 SNAP recipients (in percent). One-hundred counties (Alamance, NC to Yancey, NC, in alphabetical order) were in the 13.29% median. About 1,530 counties had fewer residents benefitting from SNAP than the median and 1,511 counties had more than the median. Twenty-three counties had the least residents benefitting from SNAP (5.67%; Albany, WY to Weston, WY), whereas 33 countries had the most residents benefitting from SNAP (22.06%; Bernalillo, NM to Valencia, NM). To simplify classification, counties in the median and with less than the median were treated as 0, and counties with more than the median were treated as 1.

The random forest algorithms built 160 trees, with some misclassifications: 29 observed less but predicted more and 28 were mismatches that observed more but predicted less, against 18,586 (both less) and 22,523 (both more) correct classes/cases. Standard errors of both training and test data were minuscule—less than 0.1%. Our primary interest was predictor importance in classification standards. The top four predictors were use of the National School Lunch Program (NSLP), adult obesity rate, soda sales tax (retail stores), and SNAP benefits (USD per capita). Albeit significant, SNAP benefits (USD per capita) ranked lower than the aforementioned predictors. Ironically, use of the NSLP was proportionate between SNAP-benefitted counties and adult obesity rate (Soda sales tax comes next).

The algorithms also identified the least essential predictors, which were persistent child poverty, population loss, and persistent poverty. The lower each of these predictors was, the more counties there were with fewer SNAP recipients.

Conclusions/Implications

The U.S. is not free from food insecurity. Although the national SNAP participation rate changes constantly based on economic prosperity and recession, about 12% of U.S. families stand food insecure (Economic Research Service, 2022) because they lack the adequate nutrition required to live healthy, productive lives. The predictors we identified characterize counties home to more SNAP beneficiaries. The most important predictor was the NSLP, which provides low-cost or free, nutritious lunches to public and nonprofit school students. Newman and Scherpf (2013) similarly identified NSLP participation as an essential predictor of SNAP participation (Newman & Scherpf, 2013). Our findings reaffirm previous findings characterizing SNAP participants and provide a greater understanding of local/county needs that can help inform tailored resource allocation and program implementation. As agricultural educators and Extension professionals, we should continuously re-evaluate county-level SNAP classifications to determine how to educate best and serve food-insecure households and communities.

Reference

Ahn, J., Briers, G., Baker, M., Price, E., Sohoulande Djebou, D. C., Strong, R., Piña, M., & Kibriya, S. (2022). Food security and agricultural challenges in West-African rural communities: A machine learning analysis. *International Journal of Food Properties*, *25*(1), 827-844. https://doi.org/10.1080/10942912.2022.2066124

Economic Research Service. (2021, January 8). *Food environment atlas: Access and proximity to grocery store American Indian or Alaska Native, low access to store, 2015*. [Data set]. Data Planet™ Statistical Datasets: A SAGE Publishing Resource from United States Department of Agriculture. https://doi.org/10.6068/DP1863709995593

Economic Research Service. (2022, October). *Food security and nutrition assistance*. U.S. Department of Agriculture. https://www.ers.usda.gov/data-products/ag-and-food-statistics-charting-the-essentials/food-security-and-nutrition-assistance/

Food and Nutrition Service. (2023, January). *Food National level annual summary: Participation and costs, 1969-2022*. U.S. Department of Agriculture. https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap

Newman, C., & Scherpf, E. (2013). *Supplemental Nutrition Assistance Program (SNAP) access at the state and county levels: Evidence from Texas SNAP administrative records and the American Community Survey*. U.S. Department of Agriculture, Economic Research Service. https://ageconsearch.umn.edu/record/262218

Pinard, C. A., Bertmann, F. M., Byker Shanks, C., Schober, D. J., Smith, T. M., Carpenter, L. C., & Yaroch, A. L. (2017). What factors influence SNAP participation? Literature reflecting enrollment in food assistance programs from a social and behavioral science perspective. *Journal of Hunger & Environmental Nutrition*, *12*(2), 151-168. https://doi.org/10.1080/19320248.2016.1146194

Pugliese, R., Regondi, S., & Marini, R. (2021). Machine learning-based approach: Global trends, research directions, and regulatory standpoints. *Data Science and Management*, *4*, 19-29. https://doi.org/10.1016/j.dsm.2021.12.002

Roberts, T. G., Harder, A., & Brashears, M. T. (2016). *American Association for Agricultural Education national research agenda: 2016-2020*. Department of Agricultural Education and Communication, University of Florida. http://aaaeonline.org/resources/Documents/AAAE_National_Research_Agenda_2016-2020.pdf

Snap to Health. (n.d.). *The history of SNAP*. https://www.snaptohealth.org/snap/the-history-of-snap

The Data Planet. (2023, January). *Data Planet statistical datasets: About Data Planet statistical datasets*. SAGE Publications. https://data-planet.libguides.com/c.php?g=398594&p=2710212

United States Census Bureau. *(2022, October 10). Supplemental Nutrition Assistance Program (SNAP) - annual: Snap benefits recipients - percentage of population, 1989 - 2019*. [Data set]. Data Planet™ Statistical Datasets: A SAGE Publishing Resource. https://doi.org/10.6068/DP18636F4422387